DOCUMENT RESUME

ED 192 339 CS 205 853

AUTHOR Rose, Andrew M.; Cox, Louis A., Jr.
TITLE Following Instructions. Technical Report No. 4.
INSTITUTION American Institutes for Research, Washington, D.C.; Carnegie-Mellon Univ., Pittsburgh, Pa.; Siegel & Gale, Inc., New York, N.Y.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO AIR-75002-2/80-TR
PUB DATE Feb 80
CONTRACT 400-78-0043
NOTE 38p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adults; Functional Literacy; *Functional Reading; Language Research; *Readability; Reading Comprehension; *Reading Research; *Sentence Structure; Syntax; *Writing Skills
IDENTIFIERS Conditionals; *Document Design Project; *Instructions

ABSTRACT

Twenty-four adults participated in an experiment to determine the difficulty of conditional sentences of the kinds frequently found in the instructions in government forms. The stimulus materials were 128 sentence frames of the form, "If you are X, press button Y," where X was replaced by sixteen different coordinate structures. These sentences functioned as conditional instructions, with the antecedent "if clause" expressing a reader-referenced statement on whose truth value the command was contingent. Responses were timed from sentence onset to response, and response accuracy was recorded. Results indicated that (1) whenever possible, negated or embedded clauses should be avoided in the design of instructions (e.g. "not male and not married" should be changed to "female and single"); (2) performance in following instructions varied with the type of construction used; (3) conditional instructions of the form, "If you are not A, and are B, then do X" should be rewritten as "If you are B and not A, then do X"; (4) strings of "ors" are more difficult than strings of "ands"; and (5) response time for wrong items was substantially longer than for correct items. (MKM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED192339

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AIR-75002-2/80-TR

DOCUMENT DESIGN PROJECT

Technical Report No. 4

Following Instructions

Andrew M. Rose
Louis A. Cox, Jr.

A product of the Document Design Project funded by
the National Institute of Education, DHEW, Washington, DC
Contract No. NIE-400-78-0043

February 1980

C S 205P503

AMERICAN INSTITUTES FOR RESEARCH/1055 Thomas Jefferson Street, NW, Washington, DC 20007 • 202/342-5000
with Carnegie Mellon University and Siegel & Gale, Inc.

OCT 2 0 1980

This report is part of the work being done under contract #400-78-0043 of the National Institute of Education. The project officer is Dr. Ramsay Selden of the Teaching and Learning/ Reading and Language Group. This report does not necessarily reflect the views of the funding agency.

ABOUT THE DOCUMENT DESIGN PROJECT

In September 1978, the American Institutes for Research (AIR) began the Document Design Project to foster clear and simple writing and design of public documents. The purpose of the Document Design Project (DDP) is to help make forms, regulations, brochures, and other written materials easier for people to read, to understand, and to use.

Carnegie-Mellon University and Siegel & Gale, Inc., a private firm that specializes in language simplification and language training, work with AIR on this project. Funding for the project comes from the National Institute of Education.

The Project's goal is to increase the knowledge and skills of people who produce public documents. To accomplish this goal, the Document Design Project is working on three tasks:

Task 1: To conduct theoretical and applied research studies on language comprehension, on the ways in which skilled and unskilled writers work, on problems associated with different document features.

Task 2: To work directly with government and private agencies as they produce materials for public use.

Task 3: To develop courses on writing and design for graduate students and undergraduates.

If you have questions or comments on this report or on other work of the Document Design Project, contact Dr. Janice C. Redish, Director, The Document Design Project, AIR, (202) 342-5071.

Table of Contents

INTRODUCTION 1

BACKGROUND 2

Uses of Conditionals 2

Previous Research on the Problem 4

Previous Research on Sentence Verification Tasks . . 6

THE STIMULUS SET 8

METHOD 12

Subjects 13

Materials and Apparatus 13

Procedure 13

RESULTS 15

"And" vs. "Or" 19

Logical Equivalents 23

Symmetric Equivalents 23

Ordering of Sentence Types 25

Dual Equivalences 30

CONCLUSIONS AND IMPLICATIONS 31

LIST OF TABLES:

Table 1: Summary of ANOVAs 16

Table 2: Descriptive Information for Sentence Types. 17

Table 3: A Comparison of Symmetrically Equivalent Phrases 25

Table 4: Comparison of Models 29

LIST OF FIGURES:

Figure 1: Mean Response Time and Proportion Errors as a Function of Repetition 18

Figure 2: Mean Response Time and Proportion Errors as a Function of AND/OR 20

Figure 3: Mean Response Time for AND vs. OR as a Function of Sentence Type 22

Figure 4: Mean Response Time for Logical Equivalents 24

REFERENCES 33

FOLLOWING INSTRUCTIONS

The documents and forms which our society uses to manage its affairs are a source of serious problems. Sometimes these forms and instructions cause frustration, actual financial loss, or the loss of important opportunities. The hundreds of millions of hours that taxpayers spend preparing IRS forms and the $500 million they spend paying tax preparers are obvious examples of some costs presumed to be attributable to complicated instructions. And in spite of the time and money spent on preparing tax forms, the IRS still finds that taxpayers make over three million math errors a year.

In this paper, we consider one particular type of instruction: conditional ("if. . ., then . . .") statements. Conditionals occur frequently on forms, in instructions for filling out forms, in operating manuals, and in many other contexts. For example, the first page of the instructions for the IRS Form 1040 offers a steady barrage of conditionals. There are 22 "ifs" on the first page alone. The form cannot be completed unless the reader understands and correctly answers each of these conditionals.

Let us consider some possible causes of difficulty with conditional instructions. The following example is taken from a Department of Justice (Immigration and Naturalization Service) form:

> If you are the spouse or unmarried minor child of a person who has been granted preference classification by the Immigration and Naturalization Service or has applied for preference classification, and you are claiming the same preference classification, or if you are claiming special immigrant classification as the spouse or unmarried child of a minister of religion who has been accorded or is seeking classification as a special immigrant, submit the following . . . (INS Form I-485).

Even the sophisticated reader has a hard time working through this sentence. There are three potential sources of

difficulty: First, the meaning of specific words or phrases, such as "spouse," "accorded," or "preference classification," might be unknown (especially to immigrants). Second, even if the words are understood, the reader might have difficulty deciding to which category he/she belongs (e.g., "Am I an unmarried child of a minister of religion who has been accorded classification as a special immigrant?"). Third, the logical structure (i.e., the sequence of "ands" and "ors") might be difficult to unravel.

This paper reports on the results of a study exploring the third problem--difficulty with the logical structure of conditionals. The study has three functions:

- It identifies a set of theoretical processes involved in understanding conditionals;
- It provides empirical data on relative response time and accuracy for a subset of all possible two-category conditionals;
- It provides practical evidence for the relative benefit of alternate presentations of conditionals.

Background

Uses of Conditionals

The most common form of a conditional instruction is the "imperative conditional."* The basic form is as follows:

"If you are in state X, then take action U."

*Fillenbaum (1977, 1978) and others have identified several other types of conditionals. For example, the following four types are common:

1. The material conditional: "If it is true that P, then it is true that Q." (P and Q are statements, each being true or false.)
2. The generalized conditional: "If anything is a P, then it is a Q." (P and Q are names of classes.)
3. The temporal causal conditional: "If you do P, then Q will happen." (P and Q are actions.)
4. The contrafactual conditional: "If it were true that P, then it would be true that Q." (P and Q are statements.)

The antecedent (the "if" clause) typically specifies or defines particular categories; the consequent (the "then" clause) typically directs the reader to perform some act. The imperative conditional is a refined "go to" statement, telling *under what condition(s)* the instruction is to be performed.

Imperative conditionals are the mainstay of federal forms. As examples, consider the following from a tax form: "If Form 1040A, line 7, is $4,000 or more, enter $400 here;" or "If Form 1040A, line 10, is $4,000 or less, do not complete the rest of this worksheet." Even the Commissioner's concluding comment, "If you need help, please call us . . ." is an imperative conditional.

Imperative conditionals play two important roles in the design of forms. At the global organizational level, they are used to *direct processing flow*; that is, they tell the reader where to go next on the form itself. For example, page 26 of the 1978 Form 1040 includes the following imperative conditional: "If you need more information about business income and expenses, get Publication 334, Tax Guide for Small Businesses." Similarly, imperative conditionals are used to terminate processing, as in "If line 7 is $4,000 or less, do not complete the rest of this worksheet." (Form 1040, p. 2, 1978)

At the level of individual items, imperative conditionals are used to *elicit responses* (usually information) from the reader; that is, they tell the reader what to enter on the form. For example, item 4 of the Earned Income Credit Worksheet (p. 2 of the 1978 1040) begins, "If you were self-employed, enter the amount, even if a loss, from Schedule SE, line 13." Item 8 includes the instruction "If line 7 is more than $4,000, complete lines 9 through 13." Item 4 is an archtypical response-eliciting use of the imperative conditional ("what to do" type); item 8 directs processing flow.

To reiterate, the imperative conditional plays two important roles in handling the information flow required to complete a form: (1) *It tells the user where to go*, based on conditions which vary from user to user (e.g., age, sex, or income). Thus, it directs the processing flow in a way which is sensitive to differences between users. (2) *It tells the users what to do* when they get there (e.g., entering certain data on the lines provided). Again, it selects the response demanded on the basis of distinctions between users. Thus, the imperative conditional is used to make a form or document sensitive to particular features in its audience.

Previous Research on the Problem

There are several lines of evidence which indicate that conditionals are hard to understand--often disproportionately so with respect to the number of words in them. The study of formal logic reveals that very complex relationships can be expressed in relatively few words using conditionals. In essence, a conditional statement is the representation of all combinations of the truth values of the categories; thus, the simple statement "If you are X and Y, do Z" in fact substitutes for four separate statements:

If X is true and Y is true, do Z.
If X is true and Y is false, don't do Z.
If X is false and Y is true, don't do Z.
If X is false and Y is false, don't do Z.

Some empirical evidence showing that conditionals are difficult to understand was obtained from the IRS, which maintains line-by-line error rates. The largest proportion of errors occurs on line 33, IRS Form 1040. The IRS also receives far more questions and complaints from the public on this item than on any other item. Item 33 is basically one long conditional, with periods and headings inserted. It starts off (after a seven-line preamble) with:

You must itemize deductions if:

a. You can be claimed as a dependent on your parents' return and had interest, dividends, or other unearned income of $750 or more.

It concludes with:

> If your itemized deductions are less than the amount shown above for your filing status (or you choose not to itemize), enter zero on line 33 unless you must itemize as described above.

Finally, the difficulty which people experience even with fairly simple conditionals has been well documented in the psychological literature. Wason (1966) has shown that even highly intelligent subjects find it extremely difficult to understand the "simple" conditional, "If a card has a vowel on one side, then it has an even number on the other side." "Understanding," in this case, is determined by asking which of the four cards

A D 4 7

must be turned over in order to see whether or not the conditional is true for all four cards. (More than one card may be turned over, but only the one or ones which *necessarily* must be examined are permitted.)

The most common responses (Johnson-Laird & Wason, 1977) are "A and 4," and "only A."

These are both wrong, and show that the conditional, despite its simplicity, is usually misunderstood. (The correct answer is A and 7.)

Other psychologists (e.g., Sternberg, 1977, 1978) have documented the relative difficulty of conditionals, considered in the context of rule-learning and concept-learning paradigms. Early research on the acquisition of bi-dimensional (i.e., two-category) conceptual rules with specifically constructed stimulus referents has demonstrated a reliable order of rule difficulty (from easy to hard): simple affirmative ("the item is square"), conjunctive ("the item is square and red"), disjunctive

("the item is square or red"), conditional ("if the item is square, then it is red"), and bi-conditional ("the item is square if and only if it is red"). (See Neisser & Weene, 1962; Haygood & Bourne, 1969.)

The conjunctive-disjunctive ("and"/"or") difference persists in the wider research on classification processes. Subjects show consistent biases toward selecting, and greater facility in learning, conjunctive over disjunctive classifications (e.g., Bruner, Goodnow, and Austin, 1956; Hunt and Hovland, 1960; see also Braine, 1976, for the precedence of "and" over "or" in the speech of children).

However, the work on rule-learning is, for the most part, unconcerned with the comprehension and *use* of sentences expressing conceptual rules. The typical experimental task presents the subject with a succession of stimulus items which are instances of the rule, and requires the subject to demonstrate concept attainment by consistently identifying further instances. Our concern, on the other hand, is not with acquisition but with interpretation and application of rules. We would expect, however, that the order of difficulty found in the acquisition tasks should generalize to application tasks.

Previous Research on Sentence Verification Tasks

The typical paradigm used to investigate sentence comprehension is the sentence verification task. Usually, performance on this task requires subjects (Ss) to assess the truth of compound statements. These statements are of the form "If you are X, do Y," where "X" is the compound of two categories. It is this compound (the antecedent clause) which contains the independent variables of interest. Studies designed to reveal differences in difficulty and the sources of those difficulties within the logical structure typically look at the variables that define logical meaning. These studies usually restrict themselves to statements with compound categories (or propositions)

joined in coordinate relationships--that is, categories linked by "and" and "or," rather than by subordinate relations (such as "because" and "when"). Standard logic (e.g., Quine, 1965) calls such compounds "truth functions" because the truth value of any one of them can be determined purely from the truth values of its component categories, regardless of the meaning of those categories.

Three elementary operations are recognized by standard logic in making truth functional compounds: negation, conjunctions, and alternation. These operations correspond to "not," "and," and "or" in natural language. "And" and "or" form a complementary pair: they are the alternate ways of connecting one coordinate category (proposition) with another. This analysis gives us two basic dimensions: affirmative-negative and "and"/"or."

Previous research indicates that negative sentences ("It is not A") are harder than affirmatives ("It is A") in most instances (see summaries of the evidence in Wason, 1962, and Trabasso, 1972). Although the difficulty disappears in certain contrasts (e.g., plausible denial, Wason, 1965; change in meaning, Greene, 1970; and other situations that are natural to negation, Wales & Grieve, 1969), the general conclusion holds for the majority of experimental tasks.

How negation varies with other factors has been studied to a limited extent. The *scope* of negation (the size of the unit modified) has been found to affect the complexity of negative sentences (Carpenter & Just, 1975), as has the sentence *truth value*. The latter effect is well replicated (see citations in the Carpenter & Just review, 1975): specifically, false negatives are usually easier than true--the reverse of the order for affirmative sentences.* *Position* of the negative word in the sentence has also been varied but with equivocal results (Jones, 1968; Wason, 1970; Carpenter & Just, 1975).

*But see Trabasso, et al., 1971, for evidence that the true-false effect depends on whether sentence or picture is presented first in the verification task.

The number of negative words in a sentence has been looked at indirectly, and the results are tentative. Not surprisingly, adding negations appears to create successive comprehension decrements (Sherman, 1976; Charrow & Charrow, 1978, 1979; Johnson-Laird, 1970; Legrenzi, 1970). But critical to the number of negative words in a multiply-negated sentence is the structural relationship between those words: whether they relate independently (e.g., "not X and not Y") or hierarchically (one embedded within the scope of another, as "not both X and not Y"). Johnson-Laird, using double negation of the hierarchical type ("No X are not Y"), found that such structures severely impair comprehension unless subjects adopt a translation strategy, recoding the two negatives as affirmatives.

But amount and structure of negation have not been systematically manipulated. Finally, the effect of these variables on errors in sentence verification and other comprehension tasks has not been probed, the traditional negation studies having tested primarily reaction time. It is with these unexplored areas that we are concerned here.

The Stimulus Set

Our original intention was to include as stimuli all combinations of the three operators ("and," "or," and negation) for two-category statements. If we call the two categories A and B, it is straightforward to generate an exhaustive set (which includes only two categories):

Ands	Ors
A and B	A or B
A and not B	A or not B
not A, and B	not A, or B
not A, and not B	not A, or not B
B and A	B or A
B and not A	B or not A
not B, and A	not B, or A
not B, and not A	not B, or not A

It can be seen that this set was generated by connecting all combinations of positive and negative categories with "and" and "or." Notice also that the bottom half of the table is equivalent to the top half, since "and" and "or" are commutative. Typically, this bottom subset of possible stimuli has been ignored in the concept-learning and rule-learning literature for an obvious reason: using geometric (or pictorial) stimuli, the symmetric pairs appear as the same stimuli. That is, if a triangle is red and large, it is also large and red. However, it is not altogether obvious that semantically they are equivalent. Let us put this problem in the context of filling out instructions: There is no logical difference between the statements "If you are male and married, do X" and "If you are married and male, do X." If one statement were shown to be more difficult than the other, it could only be due to the configuration of truth values for the reader, or to some presuppositional semantic interaction between the words themselves. This type of difference would not be generalizable; and, in the absence of such information, we would have no a priori reason to expect one version to be more difficult than the other. In practice, we would have no reason to prefer one version to the other in an instruction.

However, in some cases the "symmetric" exchange could lead to different expressions of the underlying meaning. For example, consider the pairs "not A or B;" and "B or not A." In this case, we might expect the latter statement to take less time than the former, due to what we shall call "pseudoambiguity."

Consider the imperative conditional "If you are not single or female, fill out form G." If we equate "single" to "A" and "female" to "B," we see that the antecedent in this example is of the form "not A or B." Without further punctuation, it is impossible to tell whether "not A or B" or "not A or not B" is the correct interpretation of this antecedent. For example, do females fill out the form? If the first interpretation is

the case, they do; if the second, they don't. The pseudoambiguity hypothesis assumes that even with punctuation, some residual confusion between the two possible interpretations remains, and that the necessity of figuring out the correct one makes a sentence with two potential meanings (ignoring punctuation) more difficult to decipher than a sentence with only one potential meaning. The distinction between "not A and B" and "B and not A" is that the latter has only one possible interpretation, while the former could be confused with "not both A and B." Correct use of punctuation, as in "not A, and B," *formally* disambiguates the antecedent, but in practice it may well *seem* ambiguous at first glance.

A different problem arises when it is realized that another set of 16 expressions could be generated by negating each of the existing expressions (e.g., "not both A and B"). However, similar to the symmetric equivalents, this set has not been studied. This is due to the principle of *logical duality* or "the law of negative converse." This rule states that any logical expression will retain its meaning if "ands" are replaced by "ors," "ors" by "ands," and all parts negated. Applying this rule to the top half of the set of stimuli, we can create the following set of logical equivalents (using the symbols "·" to stand for "and," "v" to stand for "or," and "-" to stand for "not"):

Original	Logical Equivalent
A and B (A·B)	not either not A or not B -(-A v -B)
A and not B (A·-B)	not either not A or B -(-A v B)
not A and B (-A·B)	not either A or not B -(A v -B)
not A and not B (-A·-B)	not either A or B -(A v B)
A or B (A v B)	not both not A and not B -(-A·-B)
A or not B (A v -B)	not both not A and B -(-A·B)
not A or B (-A v B)	not both A and not B -(A·-B)
not A or not B (-A v -B)	not both A and B -(A·B)

Finally, it is possible to create an infinite number of expressions by concatenating the original set (that is, combining existing expressions with "ands" and "ors"). However, it can be demonstrated that most of these combinations are either inconsistent (e.g., "A and B, and A and not B"), or redundant (e.g., "A and B, and B and A"). The only combinations that have unique meanings are: A or B, and not both A and B $(A \vee B) \cdot -(A \cdot B)$; and not both not B and A, and not both not A and B $-(-B \cdot A) \cdot -(-A \cdot B)$.

The first of these two is known as the "exclusive or": A or 3, but not both. The second expression is the "biconditional": "A if and only if B" or "A is equivalent to B." (Incidentally, the negation of the "exclusive or" is equivalent to the biconditional and vice versa.) We have chosen not to study either of these expressions, mainly because they involve semantic expressions that might not be normally understood (e.g., "if and only if"); also, alternate rewordings would involve excessively long statements. Finally, previous evidence (e.g., Neisser & Weene, 1962) has shown that these expressions are unquestionably more difficult than any other expression thus far generated.

To summarize, this study is concerned with several aspects of imperative conditionals. First, we hoped to establish the relative difficulty (in terms of time and errors) for different combinations of variables in a sentence verification paradigm. In the course of this investigation, some theoretical formulations the predict the ordering of types were evaluated. Second, we examined the relative difficulty of the symmetric equivalents and the logical equivalents. Third, we examined the effects of "ands," "ors," and "nots" and the effect of repetition for each sentence type.

Method

The final stimulus set in this study included the following 16 expressions:

	Stimulus	Symmetric Equivalent	Logical Equivalent
1.	A	--	--
2.	not A (-A)	--	--
3.	A and B (A·B)	--	not included
4.	A or B (A v B)	--	not included
5.	A and not B (A·-B)	7	16
6.	A or not B (A v -B)	8	15
7.	not A, and B (-A·B)	(5)	14
8.	not A, or B (-A v B)	(6)	13
9.	not A, and not B (-A·-B)	--	12
10.	not A, or not B (-A v -B)	--	11
11.	not both A and B -(A·B)	--	(10)
12.	not either A or B -(A v B)	--	(9)
13.	not both A and not B -(A·-B)	15	(8)
14.	not either A or not B -(A v -B)	14	(7)
15.	not both not A and B -(-A·B)	(13)	(6)
16.	not either not A or B -(-A v B)	(14)	(5)

"A" alone (in "A" and "-A") was replaced by each of the words (single, married, male, female). In each type involving both A and B, A and B were replaced by the eight possible pairs of these same words (single-male, single-female, married-male, married-female, and the reverse of each pair). Thus, there were eight "replications" of each of the sixteen types, for a total of 128 stimuli.

Subjects

Twenty-four adults participated in the experiment. All were native English speakers. About half of the Ss were students recruited from local universities, while the remaining were volunteers from a community service foundation (located in the same building as AIR). The 24 Ss were selected so that an equal number (N=6) were single females, single males, married females, and married males.

Materials and Apparatus

The stimulus materials were 128 sentence frames of the form, "If you are X, press (button) Y," where X was replaced by each coordinate structure described above. These sentences functioned as conditional instructions, with the antecedent ("if" clause) expressing a reader-referenced statement on whose truth value the command ("do" clause) was contingent.

The experiment was conducted on an HP3000 computer, using a 2640B CRT to display the sentences. The terminal's keyboard was modified slightly to serve as the response panel. As it appeared to the S, the keyboard had three buttons, marked "A," "B," and "NA." The rest of the keyboard was covered. The three buttons were about 1.5" equidistant from a center point marked "Start," where S kept the response finger poised between trials.

In addition to presenting the stimuli, recording responses, and controlling the timing of events during the experiment, the computer also randomized presentation order of the 128 sentences with the following restrictions: each sentence type (of the 16) occurred four times in each half of the experiment, and no sentence type occurred more than twice in consecutive trials.

Procedure

Ss were tested individually in a single session lasting 35-45 minutes. Each S responded to all 128 instructions. Ss sat at the terminal facing the CRT screen. They indicated their

responses by pressing the appropriate key; the index finger of the preferred hand was used throughout.

Each session began by familiarizing Ss with the purpose of the experiment and the apparatus. They were instructed to respond as rapidly and as accurately as possible. Before the actual test trials began, Ss completed eight practice trials in order to acquaint them with the procedures and to provide a warmup. Feedback regarding accuracy was given for these practice trials (but not during the test trials).

Each trial began with the word "READY" flashed on the center of the display screen for two seconds. Immediately after the offset of the signal, Ss saw one of the sentences with its initial letter near the left screen margin. The sentence remained on the screen until a response was made, which terminated the trial. If there was no response within 60 seconds, the trial was terminated. A three-second interval separated the trials. The test trials were presented in blocks of 32, with a variable rest period between blocks.

Responses were timed in hundredths of a second from sentence onset to response. Response accuracy was also recorded. This was possible because the computer program contained a record of the correct response for each combination of subject characteristics for each S. These values were entered at the beginning of the experimental session.

Results

Analyses of variance (ANOVAs) were conducted for the response time and proportion correct measures.* Each analysis was a 16 (sentence type) x 8 (replication) design. Results are shown in Table 1. As can be seen, the main effect of sentence type was significant ($p < .005$) for both measures. Of course, this means simply that the sentence types were different, and implies nothing about the relative difficulty of the various types or the ordering of complexity. To get a better picture of the Ss' performance, descriptive information (means and standard deviations) for each type is shown in Table 2. The question of the ordering of sentences involves some theoretical issues; we will postpone that discussion until the remainder of the data have been presented.

The ANOVA for response time also produced a significant main effect ($p < .05$) of repetitions. This is shown graphically in Figure 1. Although this figure does not look like a typical learning curve, response times decreased over practice.** The sudden drop in response time shown in the figure is probably due to a combination of factors affecting subjects' responses, such as trial spacing, fatigue, and/or post-rest reminiscence. There was no significant reduction in errors across practice, nor were there significant interactions between sentence type and repetition.

The descriptive data also suggest a strong relationship between speed and accuracy. In fact, the correlation between

*No trial ever lasted to the cutoff time of 60 seconds. Therefore, all 24 Ss had a complete data set. The response time ANOVA was performed on log-transformed data to reduce the effect of extreme scores on the test and to reduce the effect of the correlation between response time and variance.

**It should be pointed out, however, that repetitions were randomized across Ss and within blocks of 64 trials; therefore, the first repetition for each sentence type did not necessarily occur during the first block of 16 trials. Thus, the function illustrated in Figure 1 is only an approximation of a "true" practice curve.

TABLE 1. Summary of ANOVAs

A. Response Time[1]

Source	df	MS	F	p<
sentence type (S)	15	45.77	131.8	.001
error	345	0.35		
repetition (R)	7	1.13	3.97	.001
error	161	0.29		
S x R	105	0.17	1.05	NS
error	2415	0.16		

B. Proportion Correct

Source	df	MS	F	p<
sentence type (S)	15	4.115	24.06	.005
error	345	0.171		
repetition (R)	7	0.131	1.28	NS
error	161	0.102		
S x R	105	0.101	1.10	NS
error	2415	0.915		

[1]Data were log-transformed prior to this analysis.

TABLE 2. Descriptive Information for Sentence Types

Sentence Type[2]	Mean RT (s.d.)[1]	Proportion Correct (s.d.)
1. A	2.51 (1.16)	0.990 (0.10)
2. -A	3.48 (2.18)	0.984 (0.12)
3. A·B	3.51 (2.26)	1.000 (0.00)
4. AvB	4.15 (2.24)	0.963 (0.19)
5. A·-B	4.38 (1.94)	0.979 (0.14)
6. Av-B	5.29 (2.86)	0.943 (0.23)
7. -A·B	5.91 (4.39)	0.958 (0.20)
8. -AvB	7.79 (5.68)	0.849 (0.36)
9. -A·-B	6.43 (3.99)	0.969 (0.17)
10. -Av-B	7.06 (4.28)	0.927 (0.26)
11. -(A·B)	8.89 (6.02)	0.745 (0.44)
12. -(AvB)	9.61 (6.26)	0.724 (0.45)
13. -(A·-B)	11.59 (7.55)	0.662 (0.47)
14. -(Av-B)	12.54 (8.90)	0.656 (0.48)
15. -(-A·B)	12.44 (7.91)	0.609 (0.49)
16. -(-AvB)	14.04 (9.59)	0.667 (0.47)

[1]Entries are in seconds, rounded to the nearest hundredth.

[2]Symbols used are: · = "and"; v = "or"; - = "not."

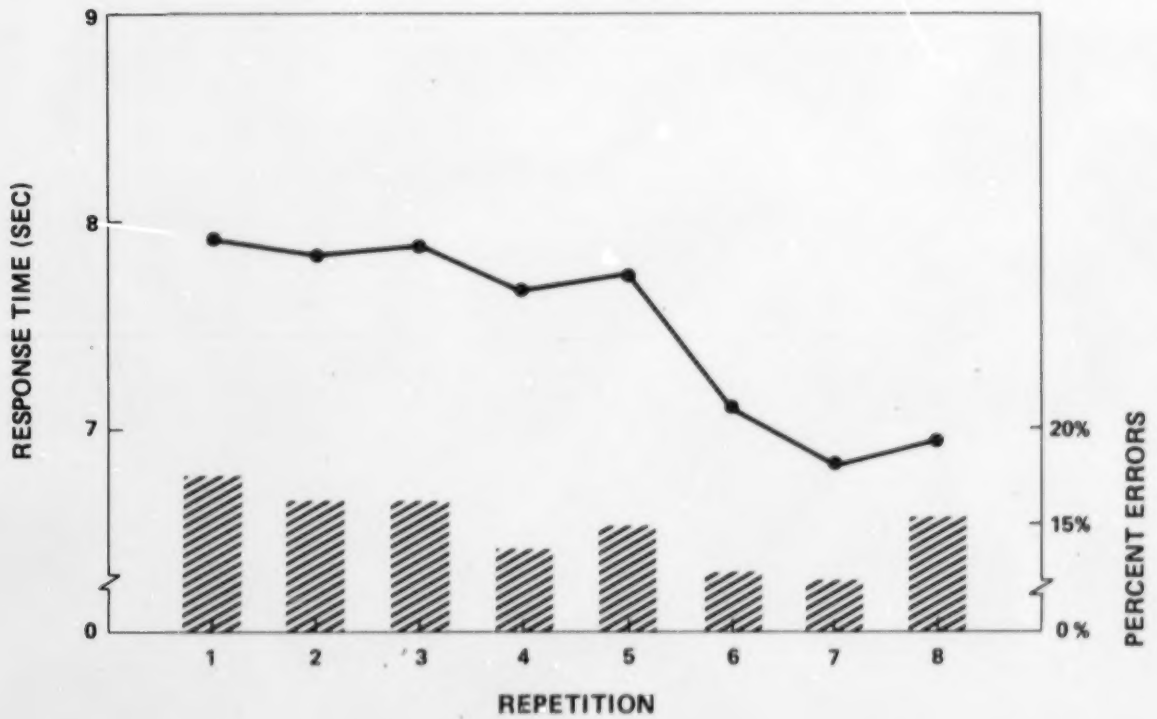

FIGURE 1. Mean Response Time and Proportion Errors as a Function of Repetition.

response time and proportion correct (calculated by correlating the mean response time and proportion correct for each sentence type) is highly significant and negative (Pearson product-moment r ($n = 16$) = -.995; $p < .001$). The same relationship was also observed within every subject: response time for wrong items was substantially longer than for correct items. This negative relationship (the slower the response, the more errors are committed) is somewhat unusual; one might expect that faster responses would tend to be incorrect more often than slower ones. It should be recalled that Ss were given instructions which did not emphasize speed or accuracy ("Respond as quickly and as accurately as you can."). The simplest explanation is just that some sentences are harder than others--despite thinking about a sentence for a relatively long time, Ss still don't understand the logical relationships, and they get the wrong answer. Another, more technical, hypothesis is that Ss adopt an implicit cutoff time. It is as if Ss decide to invest a certain amount of cognitive "effort" in each sentence; if they haven't solved the problem, they guess. It would be theoretically interesting to explore the effects of an explicit payoff (e.g., a large "bonus" for accuracy and a small "penalty" for speed) on performance.

Given this overall picture of performance, the next set of analyses examined some specific comparisons of sentence types.

"And" vs. "Or"

The overall response time and errors for sentences containing "ands" and those containing "ors" are shown in Figure 2. Since the "and-or" dimension was not completely crossed with sentence type (i.e., Types 1 and 2 contained neither), it could not be examined in the overall ANOVAs. However, it was possible to test the significance of the observed difference.* In

*Technically, we did a beta-contrast for the "and-or" variable. We used the error term from the overall ANOVA, assuming this error term to be homogeneous.

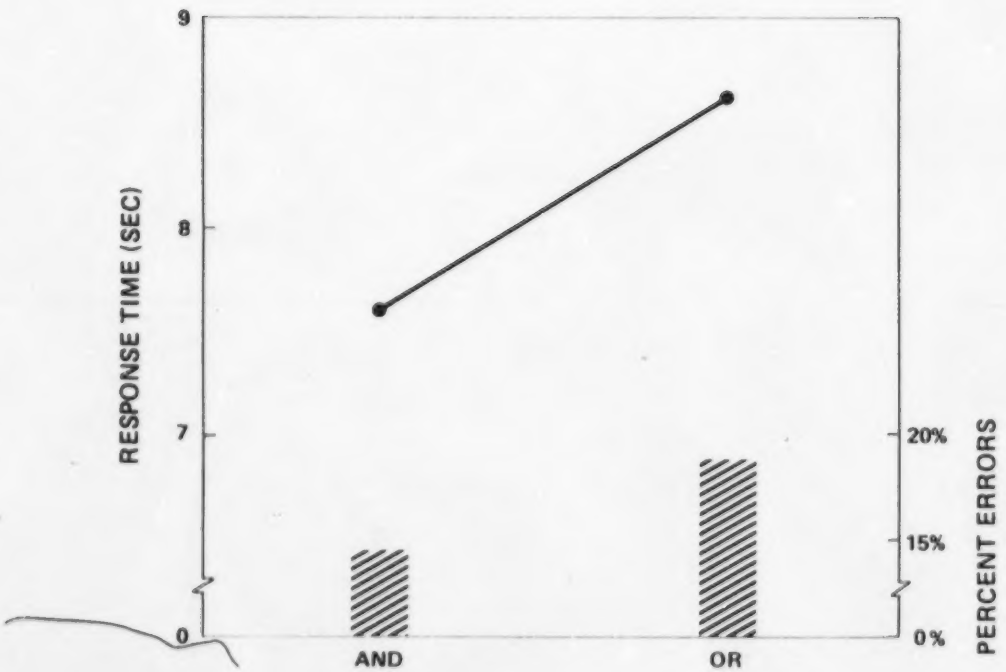

FIGURE 2. Mean Response Time and Proportion Errors as a Function of AND/OR.

fact, the difference was highly significant [$F(1, 345) = 14.88$; $p < .001$] for response time. Figure 3 shows that the "and-or" difference was obtained for every pair of sentences that were otherwise identical in construction. Thus, it is clear that "ors" take longer to process than "ands."

Although the finding that "ors" are harder than "ands" was expected, on the basis of both the literature reviewed above and the results from another experiment conducted in our lab (Holland & Rose, 1980), it is not at all clear why such a difference should occur. We can offer two highly speculative hypotheses, based on inferences concerning results obtained in the Holland and Rose study. That study used more complicated conditional statements in the same paradigm as used here (e.g., the stimulus antecedents included such sentences as, "If you are male and unemployed, or a homeowner and over 25, or not both married and retired, then . . ."). There was good evidence that Ss' initial encodings of *all* sentences were as if the connectives were "ands." If this is true, then it would take an additional step (and additional time) when the sentence was actually an "or" type.

The second hypothesis is that the conjunctive ("and") relationship is somehow more "primitive" or basic than the disjunctive ("or"). As previously cited, Braine (1976) showed that "and" typically precedes "or" in children's speech.

If this "primitive" notion is correct, "and" should be easier than "or"; it also provides a justification for the finding cited above that initial encodings of connectives seem to be as "ands." Notice also that the logical equivalent of "A or B" in "and" terms is quite complicated: "not both not A and not B." Obviously, subjects processing "or" connectives do not make this translation; however, if such a translation is made by the nervous system, it could lead to longer response times.

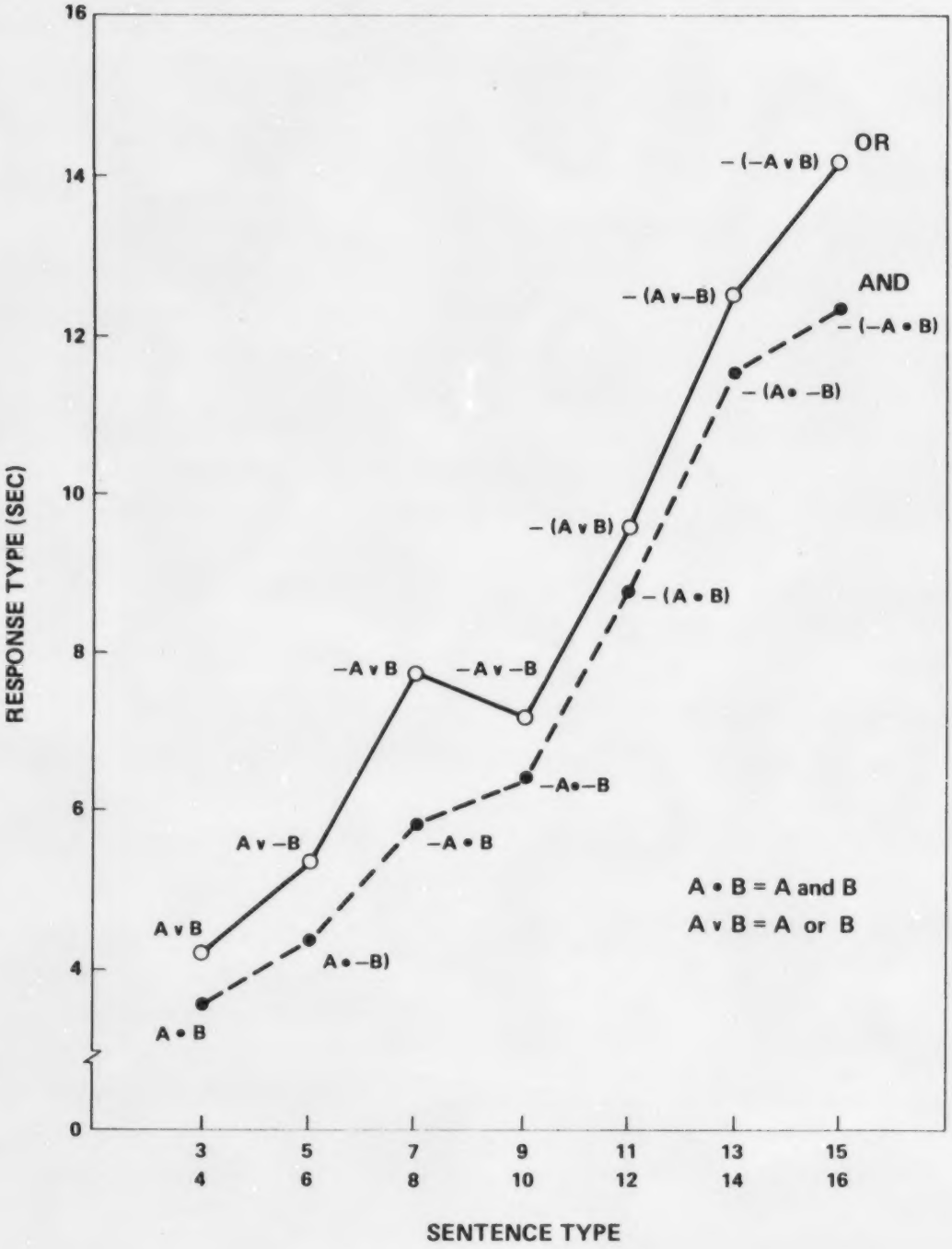

FIGURE 3. Mean Response Time for AND vs. OR as a Function of Sentence Type.

Logical Equivalents

Figure 4 shows graphically the difference between logically equivalent expressions. Clearly, Ss had difficulty with the "not both" and "not either" constructions (e.g., sentences 11-16); these constructions uniformly took longer time to process. Embedded sentences theoretically involve an additional processing step as compared to "simple" strings: "not A and not B" can be processed serially (left-to-right). However, with "not both (not A and not B)," the subject must first determine the truth value of the compound and second, transform this value.

As intuitively appealing as this argument is, however, certain comparisons create problems for a simple "number of operations" explanation. These problems will be elaborated in the Discussion section below.

Symmetric Equivalents

Table 3 below shows the comparison between pairs of symmetrically equivalent expressions. As discussed previously, these comparisons are typically ignored in sentence verification and concept learning paradigms. In accordance with our pseudo-ambiguity hypothesis, the left-hand phrases (with the "not" second) were answered more rapidly and with fewer errors than their symmetric equivalents. The differences in response time were significant (using paired t-tests) at less than the .05 level; differences in proportion correct were nonsignificant for the first and last comparisons, and significant for the middle two pairs.

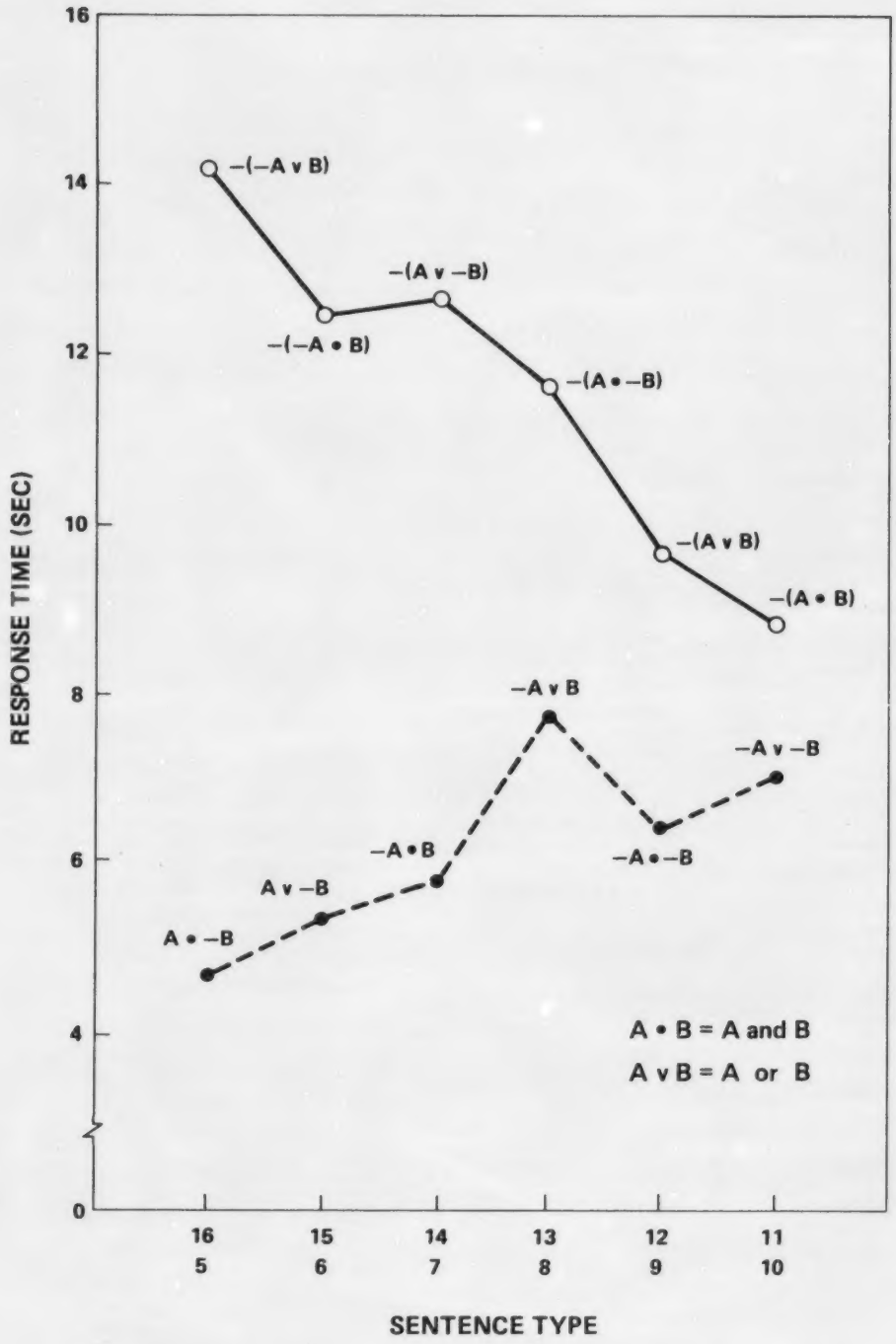

FIGURE 4. Mean Response Time for Logical Equivalents

TABLE 3. A Comparison of Symmetrically Equivalent Phrases

Phrase	mean RT	proportion correct	Symmetric Equivalent	mean RT	proportion correct
A and not B	4.38	0.979	not A and B	5.91	0.958
A or not B	5.29	0.943	not A or B	7.79	0.849
not both A and not B	11.59	0.662	not both not A and B	12.44	0.609
not either A or not B	12.54	0.656	not either not A or B	14.04	0.667

Ordering of Sentence Types

Given that these 16 sentence types differ in difficulty and that there are alternate ways of expressing each statement, a potentially useful product of this study would be a systematic explanation of the obtained ordering. For example, a "metric" of difficulty that could predict difficult constructions would be useful for document designers. Also, a theory that could account for the data would provide insights into the underlying cognitive processes involved when people unravel conditional statements.

As discussed in Haygood and Bourne (1965), there is an obvious structural difference among the unidimensional expressions (their "Level I"), expressions involving single connectives and two categories ("Level II"), and those involving two connectives (the biconditional and "exclusive or," their "Level III"). Haygood and Bourne also report that empirical evidence (e.g., Neisser & Weene, 1962) shows that the levels are reflected in concept learning difficulty. The problem is, however, that there are no hypotheses regarding differences within (specifically) Level II expressions, which make up most of the sentences used in the present study.

One hypothesis for the ordering of sentences is based on how the sentences were generated. As remarked above, in order to

generate the initial set of eight two-category concepts, we connected all combinations of positive and negative categories with "and" and "or." This systematic concept-generating rule could be stated more formally as follows: Given that A and B are the "primitive" categories upon which the set of concepts will be constructed, a complete set of concepts can be generated by applying a "not" operation and an "and" operation to each concept in an iterative fashion. Calling successive iterations "generations," this procedure works as follows: the "and" operator is applied to the concepts in the first generation, producing the concepts {A and A, A and B, B and A, B and B}. The third generation is produced in the same fashion--combining (applying the "and" operator) all previously-generated concepts, and negating all second-generation concepts. This produces the "new" concepts (after eliminating redundancies and inconsistencies such as "A and not A") of {not A and B, not B and A, not A and not B, not B and not A, not both A and B}.

This procedure continues until there are no new concepts generated. The resultant set consists of five generations and all logically possible combinations as follows:

Generation 1:	A; B
Generation 2:	not A; not B; A and B
Generation 3:	not A and B; not B and A; not A and not B; not both A and B
Generation 4:	not both not A and B; not both not B and A; not both not A and not B
Generation 5:	the biconditional and the "exclusive or"

It should be noted that "or" does not appear in this set, since it was not needed to generate all of the possible combinations (and also, the "or" equivalents are all contained in the set). It would be possible to generate the same exhaustive set of 14 concepts by using an "or" rather than an "and"; again, the procedure terminates after five generations.

A plausible hypothesis regarding sentence (or concept) difficulty is that difficulty is related to generation. That is, the harder a concept is to produce, the harder it will be to unravel. For the sentences used in this study, the prediction is that the order of observed difficulty will be (from easy to hard):

1. A
2. not A; A and B; A or B
3. A and not B; A or not B
3.5. not B and A; not B or A; not A and not B; not A or not B
4. not both A and not B; not either A or not B
4.5. not both not B and A; not either not B or A

The distinctions made between 3 and 3.5, and 4 and 4.5, are based on the "pseudoambiguity" hypothesis. Finally, we can hypothesize that further distinctions within generations are predictable based on the "and-or" difference.

In order to test these predictions, we compared the amount of variance accounted for in the original ANOVA with the proportion of variance accounted for under various assumptions. This procedure is perhaps easier to illustrate than to explain. The original response time ANOVA produced the following results:

Source	Sum of Squares	df	Mean Square	F	Probability
full model (16 sentence types)	38887.8	15	2592.5	55.03	<.001
error	16252.5	345	47.1		

The first model tested was to assume four generations and an "and-or" distinction. Thus, for example, the sentence "A and B" was coded as being in the second generation and as an "and," while "A or not B" was a third generation "or." This model produced the following:

Source	Sum of Squares	df	Mean Square	F	Probability
Model 1	34686.6	7	4955.2	105.2	<.001
Residual (lack of fit)	4201.2	8	525.2	11.15	<.001
error	16252.5	345	47.1		

The F-test shows that this model accounts for a significant proportion of the variance in the data. However, it is possible to test the residual variance (lack of fit) for significance by testing it against the error term for the full model. Thus, $F(8,345) = 11.15$, which is also significant. Therefore, this first model accounted for a significant proportion of the variance, but the remaining variance was also significant.

We tested two other models in the same manner. The first model we tried incorporated the pseudoambiguity hypothesis. Sentences with the "not A, or B" or the "not A, and B" constructions were coded as being different from the others. The second replaced the "generation" hypothesis with a "number of operations" hypothesis. This more "traditional" conceptualization is essentially a count of the hypothetical information transformations and recodings necessary to determine the truth value of the sentence. For example, consider the following equivalent pairs:

not A or not B	not both A and B
not A and not B	not either A or B

If we assign unitary weights to the operations of categorization, "and," "or," negation, and transformation, the left-hand phrases both require five operations (two categorizations, two negations, and an "and" or "or"). The right-hand phrases require only four operations (two categorizations, an "and" or "or," and a transform). Nevertheless, the latter two require more time to process. (Notice that the "and-or" difference is irrelevant; in fact, it would *add* to the processing time of the first pair of phrases as compared to the second pair of phrases. Similarly, reading time is irrelevant since all phrases contain

the same number of words.) Thus, an "additive operations" model would not predict the observed outcome (unless one wants to assume that, for example, one "transform" takes longer than two negations). Likewise, if an additive model is adopted, the basic embedded-nonembedded difference would be difficult to explain.

However, it is possible that this "operations" conceptualization, although incorrect for the above phrases, might be better than the model tested above across all sentences. Hence, we decided to test this in the same way as the other models. In this case, each operation (classification, "and," "or," negation, and transformation) was counted, yielding a value between one and five for each sentence. To summarize the hypotheses in the three models tested:

Model 1: generation, and-or
Model 2: generation, and-or, pseudoambiguity
Model 3: operations, and-or, pseudoambiguity

The results of the tests on Models 2 and 3 are shown below:

TABLE 4. Comparison of Models

Source	Sum of Squares	df	Mean Square	F	Probability
Full model	38887.8	15	2592.5	55.03	<.001
Model 2	34785.8	8	548.2	96.60	<.001
Model 2 lack of fit	4102.0	7	585.0	12.44	<.001
Model 3	34999.7	9	3888.9	82.60	<.001
Model 3 lack of fit	3888.1	6	686.0	13.76	<.001
error	16252.5	345	47.1		

It can be seen that each model accounts for successively more of the variance (i.e., sum of squares); however, given the loss of degrees of freedom (as reflected in the mean squares),

the result is that the last two models do not add any predictability over the first model.

It can be concluded from these analyses that the "generation" hypothesis, and the "and-or" hypotheses are strongly supported. However, they do not fully account for performance on this task.

Dual Equivalences

In addition to the symmetric and logical equivalences discussed in the Results section, there is another form of equivalence that results from the fact that there are one-word expressions for the negation of each category used in this study. Thus, "not male or not married" is equivalent to "female or single." While this is a trivial transform (the second is clearly easier than the former), other transforms based on logical duality (the law of negative converse) are perhaps more interesting. For example, "male or married" is logically equivalent to "not both female and single." Again, while the second is apparently more difficult, if we take a further step and negate the consequent, it might actually be easier. Thus, here are two logically equivalent versions of the same instruction:

If you are male or married, fill in line 7.
If you are female and single, don't fill in line 7.

The current experiment did not test this derivation; more generally, this sort of transform depends upon the existence of mutually exclusive and exhaustive population partitions as categories. However, the point is that there can be several alternate formulations of a conditional expression; empirical evidence could make a critical contribution to the decision to use one form or another.

Conclusions and Implications

In this report we have established a number of implicit rules and guidelines which can help the forms designer simplify the complex conditionals in forms. We also have a few practical suggestions for the designer:

- Conditional instructions of the form "If you are not A, and B, then do X" should be rewritten as "If you are B and not A, then do X." This will probably reduce both processing time and error rate. This suggestion holds for all of the "pseudoambiguous" constructions.
- Conditional instructions of the form "If you are not either not A or B, then do X" should be rewritten as "If you are not either B or not A, then do X." This suggestion holds for all of the "not both" and "not either" constructions.

In simpler terms, the suggestion is that whenever possible, negated or embedded clauses should be avoided in the design of instructions. While we have tested one form of rewrite (the logically equivalent "string" form), it is still apparent that these types of instructions cause problems for readers. We are currently testing different forms of rewrites (e.g., algorithms, and "jump"-questions) for difficult sentence types; it is too early to present specific recommendations. Probably the safest procedure to follow is to divide negated or embedded clauses into two separate sentences. While this will most likely increase processing time, it will also reduce errors. Finally, it is apparent that performance in following instructions varies with the type of construction used. Alternate (logically equivalent or symmetrically equivalent) forms are *not* the same. In order to select the best version, empirical evaluations should be conducted.

REFERENCES

Braine, M.D.S. Children's first word combinations. *Monographs of the Society for Research in Child Development*, 1976, *41*, (1, Serial No. 164).

Bruner, J.S., Goodnow, J.J., & Austin, G.A. *A study of thinking*. New York: Wiley, 1956.

Carpenter, P. & Just, M.A. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 1975, *82*, 45-73.

Charrow, R. & Charrow, V. Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review*, 1979, *79*, 1306-1374.

Charrow, V. & Charrow, R. *The comprehension of standard jury instructions: A psycholinguistic approach*. Arlington, Va.: Center for Applied Linguistics, 1978.

Fillenbaum, S. Mind your p's and q's: The role of content and context in some uses of *and*, *or* and *if*. In Bower, G.H. *The psychology of learning and motivation* (Vol. 11). New York: Academic Press, 1977.

Fillenbaum, S. How to do some things with 'if.' In Cotton, J.W. & Klatzky, R.L. (Eds.) *Semantic factors in cognition*. Hillsdale, N.J.: Lawrence Erlbaum, 1978.

Greene, J.M. The semantic function of negatives and passives. *British Journal of Psychology*, 1970, *22*, 14-27.

Haygood, R.C. & Bourne, L.E. Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review*, 1965, *72*, 175-195.

Holland, V.M. & Rose, A.M. *Understanding instructions with complex conditions*. (Technical Report No. 5). Washington, D.C.: The Document Design Project, American Institutes for Research, 1980.

Hunt, E.B. & Hovland, C.I. Order of consideration of different types of concepts. *Journal of Experimental Psychology*, 1960, *59*, 220-225.

Johnson-Laird, P.N. Linguistic complexity and insight into a deductive problem. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.), *Advances in Psycholinguistics*. Amsterdam: North Holland, 1970.

Johnson-Laird, P.N. & Wason, P.C. A theoretical analysis of insight into a reasoning task. In P.N. Johnson-Laird and P.C. Wason (Eds.), *Thinking: Readings in cognitive science.* Cambridge: Cambridge University Press, 1977, 143-150.

Jones, Sheila. Instructions, self-instructions and performance. *Quarterly Journal of Experimental Psychology*, 1968, *20*, 74-78.

Legrenzi, P. Relations between language and reasoning about deductive rules. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.), *Advances in psycholinguistics.* Amsterdam: North-Holland, 1970.

Neisser, U. & Weene, P. Hierarchies in concept attainment. *Journal of Experimental Psychology*, 1962, *64*, 640-645.

Quine, W.V.O. *Elementary logic.* New York: Harper & Row, 1965.

Sherman, M. Adjectival negation and the comprehension of multiply-negated sentences. *Journal of Verbal Learning & Verbal Behavior*, 1976, *15*, 143-157.

Sternberg, R.J. Component processes in analogical reasoning. *Psychological Review*, 1977, *84*, 353-378.

Sternberg, R.J. Isolating the components of intelligence. *Intelligence*, 1978, *2*, 117-128.

Trabasso, T., Rollins, H., & Shaughnessey, E. Storage and verification shapes in processing concepts. *Cognitive Psychology*, 1971, *2*, 231-289.

Wales, R.J. & Grieve, R. What is so difficult about negation: *Perception and Psychophysics*, 1969, *6*, 327-332.

Wason, P.C. The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 1965, *4*, 7-11.